codecademy

Calculating Churn Rates: Codeflix

Segments Really Matter

Madeline Dummerth

Table of Contents

Codeflix: Getting to Know You

Getting to Know You – How Many Months and Segments?

Upon running a simple select statement for the subscription tables you find that:

- Codeflix has been operating since December 2016 through March 2017
- The months with enough data to calculate churn are January, February, and March because Codeflix does not allow users to start and end their subscription in the same month
- There are 2 different segments in this data: 87 and 30

```
SELECT *
FROM subscriptions
LIMIT 100;

/*The table below contains a few rows
pulled out of the data for example*/
```

Id	subscription_start	subscription_end	segment
1	2016-12-01	2017-02-01	87
13	2016-12-01	NULL	30
14	2016-12-01	2017-03-07	30

Codeflix: What’s Your Churn?

2. What's Your Churn?

- Calculating churn requires dividing the number of cancellations by the number of total subscribers for each time period you are looking to learn about.
- Before that calculation can happen one must manipulate the data to sum up the users. This requires creating a table for months, a cross join of months and subscriptions, status, and status aggregate.
- The status table is created by using a series of CASE WHEN statements to create the active and canceled columns for each segment like the one below:

```
CASE WHEN
            segment = 87 AND
            (subscription_start < first_day) AND
            ((subscription_end > first_day) OR
            subscription_end IS NULL)
            THEN '1'
            ELSE '0'
            END AS is_active_87,
```

Query Results

id	month	is_active_87	is_active_30	is_canceled_87	is_canceled_30
1	2017-01-01	1	0	0	0
1	2017-02-01	0	0	1	0
1	2017-03-01	0	0	0	0
2	2017-01-01	1	0	1	0
2	2017-02-01	0	0	0	0
2	2017-03-01	0	0	0	0
3	2017-01-01	1	0	0	0
3	2017-02-01	1	0	0	0
3	2017-03-01	1	0	1	0
4	2017-01-01	1	0	0	0
4	2017-02-01	1	0	1	0
4	2017-03-01	0	0	0	0
5	2017-01-01	1	0	0	0
5	2017-02-01	1	0	0	0
5	2017-03-01	1	0	1	0
6	2017-01-01	1	0	1	0
6	2017-02-01	0	0	0	0
6	2017-03-01	0	0	0	0

2. What's Your Churn?

- The status table is useful but needs a little bit more to allow for churn calculation.
- Creating a status_aggregate table provides the sum of your original data

```
status_aggregate AS
(SELECT month,
        SUM(is_active_87) AS sum_active_87,
        SUM(is_active_30) AS sum_active_30,
        SUM(is_canceled_87) AS sum_canceled_87,
        SUM(is_canceled_30) AS sum_canceled_30
 FROM status
 GROUP BY month)
```

Query Results

month	sum_active_87	sum_active_30	sum_canceled_87	sum_canceled_
2017-01-01	278	291	70	22
2017-02-01	462	518	148	38
2017-03-01	531	716	258	84

2. What's Your Overall Churn?

- The overall churn since the company has started is approximately 22%
- To calculate the overall churn trend I added month to the select statement and added GROUP BY month to the query.
- The churn rate has increased each month.

```
(SELECT
        SUM((is_active_87) + (is_active_30)) AS
sum_active,
        SUM((is_canceled_87) + (is_canceled_30))
AS sum_canceled
 FROM status)
SELECT
          1.0 * sum_canceled/sum_active AS
churn_rate
FROM status_aggregate;
```

Query Results
churn_rate
0.221745350500715

Query Results	
month	churn_rate
2017-01-01	0.161687170474517
2017-02-01	0.189795918367347
2017-03-01	0.274258219727346

Codeflix: Segments Matter

Segments Matter

It is my conclusion that Codeflix should focus on expanding segment 30 because they have a lower churn rate. This lower rate indicates that segment 30 users are happier with their subscription than segment 87 users.

Query Results	
churn_rate_87	churn_rate_30
0.374508261211644	0.0944262295081967

Query Results		
month	churn_rate_87	churn_rate_30
2017-01-01	0.251798561151079	0.0756013745704467
2017-02-01	0.32034632034632	0.0733590733590734
2017-03-01	0.485875706214689	0.11731843575419